DOCUMENT RESUME

ED 333 748 FL 019 247

AUTHOR Griffin, Patrick E.
TITLE Monitoring Proficiency Development in Language.
PUB DATE Jul 89
NOTE 18p.; Paper presented at the Annual Meeting of the Modern Language Teachers Association of Victoria (Victoria, Australia, July 10-11, 1989). Contains small type.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Foreign Countries; *Language Proficiency; *Language Tests; Second Language Instruction; *Second Language Learning; *Skill Development; Testing; *Test Use

ABSTRACT

The Victorian (Australia) assessment approach of developing national subject profiles has potential for language assessment. Within each language area, levels of development would be identified and defined by observable language behavior, which could then be tested by a variety of test types. Assessment would be done by teachers in schools using standard assessment tasks, then interpreted according to descriptions provided for each level of proficiency. A project to develop such a test for first language proficiency consisted of workshops with classroom teachers to identify observable behaviors as criteria, intensive observation of students for validation, creation of an initial development scale, field testing of the scales, and establishment of norms. A language proficiency assessment system that similarly uses standard assessment tasks and common subject profile reporting could meet several important criteria: it would be analytical, diagnostic, and criterion-referenced; enable interpretation that is progressive, developmental, and cumulative; use consensus moderation and empirical calibration; be teacher-controlled and developed; be flexible; be reliable and valid; and describe student behavior in terms communicable to parents. What is needed is considerable development work, careful explanation to schools and community, technical assistance for districts, and resources to develop, implement, and maintain the system. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED333748

Monitoring Proficiency Development in Language.

Patrick E. Griffin
Phillip Institute of Technology

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Griffin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Annual Congress of the Modern Language Teachers Association of Victoria, held at Monash University, July 10-11, 1989.

FL019247

BEST COPY AVAILABLE

Australia is currently undergoing a awakening to the importance of the population developing proficiency in languages other than English. Our trading capacity has been recognised as being deficient when those responsible for dealing with exporters in our major trading partners cannot deal with them in their own language. Without this competence our traders are at a disadvantage. But there are more than economic reasons being espoused for the development of second language competence among the Australian population. We aspire to be a multicultural society. Our national and state educational and social justice policies outline the need for tolerance, understanding and cooperation among groups with different cultural and language backgrounds. Without access to the languages there is little possibility of gaining an understanding of the cultures and of blending the Australian community into a tolerant and cooperative society. This applies to the access of English for migrants and the access of native English speakers to the languages of our major trading and immigrant groups.

The National Policy on Languages (Lo Bianco, 1988) and the Victorian Government Languages Action Plan (Lo Bianco, 1989) outline the commitment at a government level to achieving the twin aims of developing proficiency and providing access to all to the languages and cultures of the major language groups of our migrants and trading partners. The policy and action plan clearly state that every school should offer at least one language other than English and every student should become proficient in at least one language other than the mother tongue.

> **"The goal for Australian schools is Bilingualism. That is proficiency in two languages, not necessarily equal competence but the highest level of skill possible. (Lo Bianco, 1989, p.12)**

In order to achieve this goal there are important pre conditions which include the obvious resources, sufficient teachers, appropriate curriculum, motivated school decision makers and a sympathetic school community. All of these are pre conditions to the introduction of the program and The Action Plan (Lo Bianco, 1989) outlines the general approach to achieving this. There is an additional need to chart the progress of individuals, of classes, of schools and of entire states systems towards achieving the goals of the national and state policies. The meeting of the State and Commonwealth ministers of education in 1989 have already stipulated that systems will need to gather information on the progress towards national goals. The thrust towards this approach to monitoring systems is clearly aimed at rationalizing policies. There is no point in arguing for additional resources, additional time in the curriculum or for increased status if it cannot be demonstrated how this will meet the aims and needs of policies at state and national. The terms of the policies and action plans are clear. Schools are expected to provide access to bilingualism, defined in terms of proficiency and competence, to all students. Providing instruction is not enough. Arguing for additional resources to enable the school or the system to provide the course and instruction is not enough. There are many competing for resources. There are not many competing for the chance to define the learning outcomes or to demonstrate that these can make a contribution to the realisation of the aims of the policies.

The notion of proficiency in language is essential to the development of the language curriculum. Students, classes, schools and systems all need to demonstrate their progress towards the development of proficiency. Continuous Assessment and monitoring of proficiency is considered to be central to the achievement of the goals of the policy statements. Assessment and monitoring of development of proficiency needs to be placed in the perspective of the national language policy and the desire of the Australian States and Commonwealth to monitor and profile the development of students in all areas of learning. The collective ministers of education in April this year considered the options for national assessment. Support for subject profiles, records of achievement and continuous monitoring was established. At the Australian Education Council, there was considerable support for the development of student profiles. Several a priori conditions

3 BEST COPY AVAILABLE

were established. First there had to be a close relationship between curriculum and assessment; the complexity of achievements should be reflected in assessments; and assessments should be criterion based.

Four possible approaches were considered for national assessment approaches. These were

(i) statewide testing
(ii) national testing
(iii) expert appraisal - inspectors
(iv) national subject profiles.

The preference for the Victorian approach to profiling was expressed, but it was recognised that considerable work was needed before it could be made operational. In this paper, the potential of the Victorian approach to subject profiling will be outlined for general language profiling. There is of course a great deal of work to do and most of it may need to be done through specific subject associations or with assistance from grants from state and commonwealth bodies. The profiling of languages will not be a short term approach unless much can be borrowed from other work elsewhere. It is essential however, that it be done.

Remember, it is not feasible to argue that the national languages policy enables the schools and systems to argue for more resources without the other side of the equation being put forward. Success in the development of the national languages policy does not only mean the provision of courses. It does not mean the implementation of programs of language instruction, of professional development of teachers, of the provision of materials, of support agencies. If it cannot be shown that these lead to the development of bilingualism in terms of proficiency in more than one language for all students, then the policy cannot be shown to have been successful. The input monitoring must be associated with output monitoring. There is no reason any more to believe that we can evaluate courses, programs and even systems in terms of the expenditure, the number of courses, the number of students involved, the materials used, the language laboratories developed and so on. The bottom line is the language proficiency of students who emerge from the course after exposure to the newly developed teachers, the new materials, the time given in the curriculum and the other resources put into the program. If all of this does not lead to an effective bilingual society then the program and the national language policy is a failure in terms of its primary objective.

What are subject profiles? They are really methods of reporting. Within each area of language levels, of development need to be identified for each major component. So there is a need to gain agreement on the basic components of language development and to get an agreement on what is meant by proficiency. There are numerous studies of these issues and surprisingly a great deal of agreement.

Assessment of proficiency indicates the highest level of sustained performance of an individual (Byrnes and Canale, 1989). Proficiency is defined as **observed behaviour** and cannot be accounted for by any single unitary underlying ability. There is general agreement in the language literature that proficiency develops in the four so called macro skills of speaking, listening, reading and writing. Proficiency in speaking does not imply proficiency in reading or in any other language modes. In fact the discrepancy usually exists between language modes and the discrepancy is usually higher at the more advanced levels of proficiency. Galloway (1987) argues that there are four basic areas in which criteria for assessing proficiency need to be addressed for each of the four macro skills. These are the function of the language being used, the content of the language, the context in which the language is being used and the accuracy of use. Each of these are argued to affect the way in language can be demonstrated. While these may not demonstrate the exclusive nature of language development and assessment it does present a useful framework for the assessment of language emergence and of proficiency overall.

BEST COPY AVAILABLE

There is also some need to avoid the interpretation that the four areas of assessment are also discrete. For the purposes of this discussion, it helps to simplify the frame of reference for assessment and this can be presented as follows.

	S	L	R	W
Function				
Content				
Context				
Accuracy				

So if these can be considered as the basic frame of reference then we need to identify the indicators of growth within each of these 12 areas. There is then a need to identify the levels of development such that the development of the subject profiles would need to be closely linked to curriculum development. It should then be possible to develop a framework comprising a sequence of levels through which students progress due to exposure to the curriculum. Some students of course will progress faster than others. Because the levels will refer to sequenced performance levels within a subject area they should not be directly related to age/grade performances of students. The performance of students of a particular age grade would span a number of these levels and even an individual student may be developing competence at several levels.

The levels of the profiles need to be defined observable language behaviour which is elicited by a series of assessment tasks, not unlike the Standard Assessment Tasks (SATs) currently being developed by the Victorian Curriculum and Assessment Board. The difference would need to be the identification of the levels of development in advance of the standard assessment tasks. This would avoid the now apparent difficulties of developing the assessment tasks and then interpreting what performances on these mean in terms of progress or growth in the curriculum area. The assessment tasks may have different styles in different systems or even in different schools. They could be a mix of pencil and paper tests, observation of students performances against set criteria, assignments, interviews, practical tasks, essays, reading and role playing or simulation tasks and so on. Various systems may have different preferences. A student may be assesses to be at a hypothetical level 4 if there was evidence of being able to perform the tasks that define level 4 proficiency but not the tasks appropriate to level 5 proficiency.

The following Diagram illustrates the potential of the system and perhaps what such a subject profile might look like.

Language A

Speaking				Listening	Reading	Writing
Function		Accuracy				
Context		Content				
1	1	1	1			
2	2	2	2			
3	3	3	3			
4	4	4	4			
5	5	5	5			
.	.	.	.			
.	.	.	.			

Figure 1
Subject Profile and Levels of Development

The assessments of students should be made by teachers in schools using the standard assessment tasks. The student performances can be interpreted by descriptions provided in each of the levels of the subject profiles. The levels and subject profiles can then be used to report to parents and to the ministry in turn through its reporting network which is being developed through its various branches. The ministry can then aggregate at each level to avoid school level comparisons where this is seen to be unnecessary. Moderation would be necessary to avoid localization of standards (Black, 1987) and to ensure that comparability of standards is achievable. The student assessment would be criterion based in that the achievement of the students would be described by what they can do rather than what might be expected by an age/grade group or by comparisons to other students. The figure below illustrates the relationship between the levels and the standard assessment tasks.

BEST COPY AVAILABLE

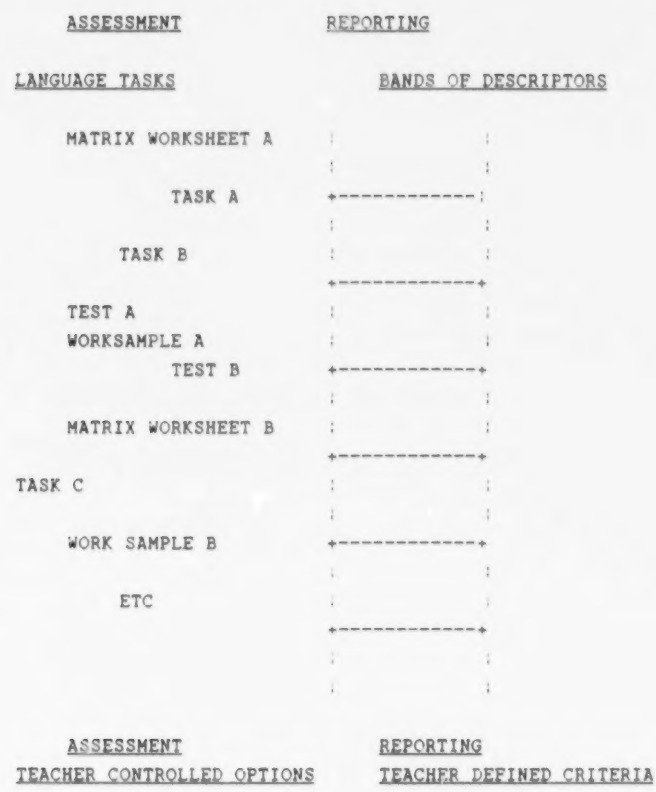


Figure 2
Posible Assessment System, Proficiency Levels and SATs.

On the left of the Figure, is a collection of potential assessment strategies such as formal tests, performance tasks (such as reading aloud), and work samples. These matrices, tests and performance tasks are expected to reflect local curriculum models. A system of moderation, across year levels and across schools can assist in developing reliability of assessments and lead to considerable common interpretation of performances.

On the right of the diagram is a symbolic representation of the proficiency levels. The levels are a reporting framework which can satisfy a number of requirements. Clearly they can be used for descriptive reporting and profiling of individual student performance and they can also be used for aggregated reporting at a system level using a rating scale method such as that presented in this paper. Norm referenced interpretation is possible where this is considered necessary. The blurring of the boundaries between criterion referenced and norm referenced interpretation is a by product of the use of item response theory.

There are several advantages of such a system of assessment which relies on standard assessment tasks and on common subject profile reporting.

Development.

The Victorian approach to profiling has been developed through the identification of reading and writing levels for students in first language. These have been based on the ASLPR and the ACTFL guidelines and the procedure used was as follows.

Method:

The study used the following steps:

(i) Workshops with classroom teachers to define the observable behaviour as indicators of development.

(ii) Intensive observations of students to validate, using group moderation, the definition of each indicator.

(iii) Surveys to identify the measurement properties of the indicators and the development of the initial development scale.

(iv) Consultation with expert informants to modify the language development scales.

(v) Field testing the scales: establishing rating norms, reliability and criterion validity estimates.

(vi) Calibrating and anchoring the band levels with specific assessment tasks.

Workshops

Almost 100 teachers spent four days spread over a school year working in syndicates of six in structured workshops, developing their skills of analysis, observation and moderation. The workshops used an analytical method which combines the **identification of goals**, the delineation of appropriate **outcomes** associated with each goal and a range of methods of gathering information, or evidence, that the outcomes have been achieved. The methods of gathering information were called **assessment methods**. These in turn were matched to outcomes for each goal. The evidence, which each assessment method was used to gather, was written into the cell of the matrix worksheet. This evidence was called the **performance indicator**.

In a series of two day workshops, the teachers were introduced to the idea of profiling using a structured program. A mixture of speakers and activity sessions both informed the teachers of the background, developments to date and expectations of the project. Using a group consensus technique (Blachford, 1985), the teachers were asked to define the areas in which language developed within the four macro skills of reading, writing listening and speaking. However, these really only help to identify more specific areas of learning. The groups of six then became syndicates for the purposes of development and remained as a working group for the duration of the project. Each syndicate was asked to define the stages of development as outcomes of learning. The teachers are asked to define the techniques of assessment they use. The difficult part was in cross referencing the results of these two sessions and creating a matrix into which the performance indicators are written. This can take a long time and typically involves a change of thinking by the participant teachers as there was a need to focus on the observable behaviour of the student and not on the interaction between the teacher and the student.

Figure 1 shows how these aspects (areas, outcomes, assessment methods and indicators) were combined into a worksheet. In the workshops teachers referred to the outcomes as milestones and the goals were referred to as **areas of literacy**. These terms were retained for the duration of the project because the teachers working in the project felt comfortable with the terms.

Place Figure 3 about here

Twenty four matrix worksheets were developed covering a range of literacy areas. An example is shown in Figure 4 below. It illustrates the relationship between the goal of **Developing an approach to Unknown Words**, the outcomes, such as:

- *Seeking help from others,*
- *Using Visual Cues,*
- *Using Auditory or grapho phonic cues,*
- *Using Semantic and syntactic cues,*

BEST COPY AVAILABLE

and the assessment methods shown as,

- *o Direct observation and anecdotal records,*
- *o Listening to oral reading,*
- *o Conferencing with students.*

Place Figure 4 about here

Note that the assessment methods offer the teacher a wide range of techniques and are in accord with the ministerial expectations of the assessment method.

Classroom Observation

Notions of Warm and Cool have been devised to assist in the field trials of the workshop materials. A warm teacher is one who has attended the workshop. A warm class is the class of the workshop teacher and a warm matrix is one developed by the teacher using it. The teachers trial their own matrices in their own classroom. That is we have a warm teacher, a warm class and a warm matrix. Clearly it is not possible to have a cold teacher with a warm matrix and this reduces the range of combinations to six. As part of the development process in workshops, four combinations were used all involving warm teachers. Later field trials involved cold teachers with cold matrices in their own (warm) classes and in other (cold) classes.

When the teachers take the matrices away and try them out in their own classroom, this is a warm trial. In these trials, the teachers see if they can recognise the performance indicators. They check to see if the description of the assessment technique is appropriate. They check to see if the milestone/outcome is a realistic description of their students' progressive development. Then they communicate these data to the workshop facilitators and with each other. They also gather examples of student work, where possible, to illustrate the performance indicator and prepare to table this at the next meeting of the syndicate at the next workshop.

At the subsequent workshop, time is devoted to discussion of the trials of the matrix in the warm classroom. Teachers have been networked between workshops as well. They had communicated any changes they saw as being necessary. Each teacher contributed their experience of observation and how the children exhibited the performance indicators. They compared their experiences and recommended revisions of the original matrix. Typical experience was that the matrix was completely re written after the warm trial.

The matrices were exchanged across syndicate groups after the initial revision and taken back to the schools for use. That is the trial involved warm teachers, using cold matrices in warm classes.

At a the next workshop day, the teachers again provided feedback on the use of the matrices in identifying appropriate student behaviour using the assessment methods in the matrix. This sets up the situation in which all teachers can comment on the work of their colleagues in developing the initial matrices and can begin to make their own revisions of others' work. This was the moderation of matrices across syndicates in that agreement had to be reached that the indicator could be observed using the assessment method and was indicative of the learning outcome listed at the top of the matrix.

Now all matrices become warm for the workshop teachers. That is, all teachers had had a hand in developing them and in revision of indicators and outcome statements. The next trial is a field test of all matrices by the workshop teachers using either a rating scale to record their observations (0=not seen;1=maybe;2=yes) or a method of recording the date on which a specific student exhibited the behaviour. The thinking behind this was to let

BEST COPY AVAILABLE

the observations of the students indicate the general trends in the patterns of emergence of language behavior. Some teachers also trialled the matrices out in classes of their colleagues at school That is we has trials using warm teachers, warm matrices and cold classes.

Only one limited trial was conducted using cold teachers, cold matrices and warm classes. The rating scale approach was used but insufficient information was given to the teachers and the use of the matrices under these circumstances was not successful. Some alternative method of presenting and training the teachers needed to be developed for this approach. The matrices were too complex, too detailed and presented the teachers with an overwhelming amount of detail and work in assessing and recording the behaviour of individual students.

Indicators and Scale Development

Instead of using the matrices, the indicators were extracted into a series of checklists. Again no teacher could possible observe all of the indicators with all of their students. There were several hundred indicators of language development. Accordingly a series of overlapping sub lists were developed so that every teacher would gather information on all indicators but it was not necessary to gather all information on every student. This ensured that every student had some observational data collected and every indicator was observed (or not). These lists of indicators were distributed among project teachers to gather data for calibration purposes. A rating scale was used to show the degree to which each of these indicators was present in the reading and writing related behaviour of the student. A zero (0) was to be used if the teacher had not observed a student exhibiting a performance indicator. A one (1) was to be used if the teacher had observed the behaviour but was not convinced that the behaviour was consistent and that this type of behaviour was still developing. A two (2) was to be used if the teacher considered that the performance indicator was now an established part of the student's repertoire of reading related behaviour. When the ratings were coupled with dates of observation of the behaviour emergence, the teachers were able to develop a short-hand way of recording their observation of the students' developing reading and writing skills. Teachers in 15 schools rated 286 students on a total of 147 indicators of reading behaviour. Teachers in 38 schools rated 578 students on 245 indicators of writing behaviour. Details of these analyses are provided by Griffin and Jones (1988) and by Griffin (1989).

The Rasch Rating Scale model of the Item Response Theory (Andrich, 1978), enabled the indicators to be calibrated so that all performance indicators could be mapped onto one continuous developmental scale. The advantage of this method is that both indicators and students can be mapped onto the same underlying growth continuum or scale. The students were then compared directly to indicators of general reading and writing development.

Proficiency Levels

The full list of indicators was examined for patterns which might be useful in summarizing the indicators into groups in similar ways to the aggregation of the indicators in the language acquisition scales such as the ASLPR (Ingram, 1984). Several patterns were evident in the list of calibrated indicators of reading behaviour. The progressions seemed to be related to underlying factors such as *attitudinal behaviour, influence of reading on writing, role playing, retelling behaviour, reactions to reading materials, analysis and interpretation, social or interactive roles in reading behaviour, word approach skills, types of reading materials used* and so on. These trends only helped to group the indictors. The labels given to them do not matter in the overall development of the proficiency scales. The groups of indicators were called bands and were developed in both reading and writing. A reading band for example, contained a description of a very broad range of reading behaviour rather than a discrete point of development. There were seven reading bands identified and nine writing bands but the number of bands does not represent anything other than the apparent groupings of indicators. The bands were

BEST COPY AVAILABLE

labeled from A through G for reading and A through I for writing, setting band A at the earliest developmental level. The bands are cumulative. That is, a student placed at Band E was likely to have the behaviour patterns indicated by Bands A, B, C and D.

Consultations with Expert Informants

The draft forms of the reading bands were distributed to teachers and a representative sample of academics, consultants, and inspectors and other ministry advisors in several Australian states, in New Zealand and in the United Kingdom. They were asked to act as "expert informants" and to review the draft version of the bands; to advise on the need to edit, delete or move the indicators included in the bands or if they considered that important indicators of the development of reading were missing, to suggest the addition and to recommend the appropriate location. Advice was also sought on the structure, appropriate use and suitability of the bands.

Field Trials

After revision by various groups of teachers and language specialists, a draft version of the Reading bands was prepared for field trial in 105 schools throughout Victoria. The writing bands were not at the same stage of development and are scheduled to be trialled in the large sample of schools in 1989. A rating scale was used which described the teachers' observations in terms of the student exhibiting

3. If the student **has established the behaviour pattern** and consistently exhibits all or most of the behaviour described in the band.

2. If the student **is developing the behaviour pattern** such that some but not all of the behaviour for a band is often exhibited, use a **code of 2** for that band.

1. If the student **is beginning to show signs of the behaviour pattern** of a band level in that only a little of the pattern is shown, use **a code of 1** for that band.

0. If the student **shows none of the behaviour pattern** for a band level, use a **code of 0** for that band.

Teachers in primary schools were asked to rate students at years 1,3 and 5, and to administer a standardised test. The Primary Reading Survey Test (Form AA) ACER, 1981) was administered to year one students and the Test of Reading Comprehension (TORCH) test (Mossensen, Hill and Masters (1984) was administered to years 3 and 5. Secondary schools were asked to rate students in years 7 and 9 and to administer the TORCH test to these students. Item level information was provided by 60 teachers and these data were used to equate the bands and the tests. The results of this analysis are reported elsewhere. Teachers in all 105 schools provided total test score and band ratings for students. These have been used to estimate the internal consistency reliability and the criterion validity. A small number of teachers were asked to rate their students before and after the school holidays in order to estimate the intra rater reliability. More than 4000 students were assessed using the reading bands.

The development of the proficiency bands has enabled two forms of monitoring to be introduced. Clearly the one-off assessment, when the data is collected using the teacher judgement as a means of assessing students, is what can be called a snapshot survey. Where the teachers were recording the dates on which they observed the behaviour emerging it is called a longitudinal approach. But the proficiency bands really combine both forms of survey. Teachers professional judgements of students work build up over a long period of time. In many instances, they are informed by such assessments as standardised tests, assignments, work samples, interviews, and other forms of assessments including student self assessment. All of this information goes into forming the teacher's judgement in a

"snapshot" application of the subject profiles. It is true that teacher judgements are affected by localisation of standards (Black, 1987) and that there are rater effects and halo effects operating. However these may be most serious at the level of the individual student and some control over them can be exercised using a system of moderation not unlike the system used by VCAB with the year 12 assessments. This would not impose any new approaches on the secondary teachers, but primary teachers involved in the project found the idea novel but valuable. The professional development spin-offs were obvious and served to assist in issues such as reliability and validity of judgements. When the teachers ratings were compared to standardised test results at a class level there was an 85% consistency between the two sets of data.

One further thing needs to be pointed out. The teachers involved in the project have developed an ownership of the scheme. This surely is a further strength of the project. Further developments of the project are planned for the future. The same approach is being adopted in ESL, Science, Social Education and Mathematics. There is Ministerial commitment to the method in Victoria; there is a general interest in using the approach in the United Kingdom through one of the consortia developing the national assessment for and there is general interest in other Australian states and in New Zealand. Expressions of interest have also been made by groups involved in language education. What is needed is a group of teachers willing to take the lead and begin the development of proficiency scales and assessment task banks for the different languages.

Advantages.

(i) The assessment system should enable an **analytical or diagnostic** approach to be adopted in assessment. The term analytical is preferred because it does not infer that there are only problems. An analysis seeks both strengths and weaknesses and provides information which can be used to identify appropriate targets and paths for teaching and learning.

(ii) The assessment system should be **criterion referenced** in that the student's performance or behaviour pattern is compared to a series of tasks. Criterion referenced interpretation enables the students development to be interpreted in terms of behaviours which they can demonstrate. If the proficiency levels contain sets of indicators which enable criterion referenced interpretation, each student's development can be interpreted in terms of the descriptive profiles of language behaviour rather than in terms of age or grade norms.

(iii) The assessment system should enable interpretation of assessment of learning to be **progressive, developmental and cumulative**. There is a need to trace out a general direction of development of students without prescribing the precise path of development for any individual student. The notion of accumulation is important. The assessment system needs to describe a progression of skills which are retained. It should neither describe behaviour in deficit terms nor in terms of transitory behaviours which might be described as stages through which students pass and leave behind. The scale needs to illustrate the general pattern of how skills accumulate without claiming to have identified all skills or to reach the definitive end of the progression. Even if an accumulation of skills can be defined, it is not true that the progress through this accumulation is always linear or monotone. For example even the most proficient reader, when placed in an entirely new context, may have to employ skills which are usually exhibited by readers at earlier levels of development. This would not mean that the proficient reader's skills have diminished. The possibility of moving forward and backward throughout the progression must be available. What is important is the idea of a threshold which is exhibited when the learner is working within familiar contexts.

BEST COPY AVAILABLE

(iv) The assessment system and the Standard Assessment Tasks should be formed using both **consensus moderation and empirical calibration**. If the proficiency levels are to be used for routine monitoring of student outcomes there may be a need for a range of forms of moderation.

(v) The assessment tasks and curriculum content should be **teacher controlled and developed**. This is important to the integrity of assessment of learning. When the teacher makes the judgements about what to assess and how it should be assessed, the information obtained is more likely to be related to the curriculum and to be interpreted in the context in which the learning occurs. Externally controlled assessments cannot always provide this direct relevance. This is not to argue that externally developed standardized tests should not be used. On the contrary, the classroom teacher should be able to identify appropriate tests which assess skills directly related to the curriculum, take account of the context and use the information with the general progression defined by the proficiency levels. A strength of the criterion levelssuch as those described by the ASLPR is that they offer a wider frame of reference for interpretation of test information than the restricted paper and pencil tasks which characterise such tests.

(vi) The use of simple rating scales with the profiency levels and the standard assessment tasks also assists in making the assessment system more **flexible**. Graded assessments can be based on recording systems which only allow complete/ incomplete or right/wrong observations of individual tasks. However, the behaviour described by each of the indicators may not be readily described as simply present or absent. Some language behaviours emerge over time and the recording process needs to allow for that if it is going to assist teachers in proper analyses of language proficiency development. Because levels such as those in the ASLPR are criterion referenced and sequences of student development, they are not directly related to age or grade levels or to expectations of students and hence relate to learning rather than specific sub groups of learners.

(vii) The assessment system needs to demonstrate **reliability and validity**. Problems associated with rater and halo effects need to be controlled as much as possible and will always be present to some extent when judgement forms the overall basis of the assessment. Face **validity** can be derived from the base of teacher development and the 'bottom up' approach.

(viii) The proposed system relies on the ability of teachers to describe behaviour of students in terms **communicable to parents**. For communication to the wider community, distributions of levels associated with the profile levels could be used based on ratings, or estimates of students at developmental stages for each band level. The progression through the levels needs to be simple to understand, and there needs to be sufficient levels to ensure that some progress is evident over a reasonable time.

There are also disadvantages.

(i) There is a need for a considerable amount of development work to be done in identifying the levels, specifying the criteria and establishing the bank of assessment tasks. This would mean that there would be considerable lead time before the approach can be fully implemented. However the trade off of the time and development effort against the benefits in terms of the professional development and curriculum pay offs should not be underestimated.

(ii) The system would have to be "sold" to schools and explained to the community- two groups which tend to be suspicious of new ideas and are resistant to change.
(iii) Some schools and teachers would need additional assistance- a consultancy service would be required. This may not be a disadvantage however, as the cross fertilisation of ideas and assessment materials would more than compensate for the effort involved.

BEST COPY AVAILABLE

(iv) Considerable resources would need to be developed to implement the system and then more to maintain it.

References.

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581-594.

Black, P. (Chair) (1987). National Curriculum : Task Group on Assessment and Testing. London: Department of Education and Science.

Blachford, K. (1985) Destinations Decisions. Melbourne. Ministry of Education.

Byrnes, H. and Canale, M. (Ed.) (1987). Defining and Developing Proficiency: Guidelines, Implementations and Concepts. National Textbook Company: Chicago.

Lo Bianco, J. (1987) National Policy on Languages. Commonwealth Department of Education, AGPS, Canberra.

Lo Bianco, J. (1989) Victoria: Languages Action Plan. Ministry of Education. Melbourne.

Galloway, V. (1987). From Defining to Developing Proficiency: A Look at the Decisions. in Byrnes, H. and Canale, M.(Ed.). Defining and Developing Proficiency: Guidelines, Implementations and Concepts. National Textbook Company: Chicago.

Griffin, P. (1989) Monitoring Literacy development: The accumulati0n of Literacy Skills. Australoian Journal of Education, (in Pres).

Griffin, P. (1988). *Literacy Profiles: A Matrix Approach.* Paper presented to the Australian Cooperative Program. Conference on Profiling, Melbourne, June 26th.

Griffin, P. and Jones, C. (1988). Assessing the development of reading behaviour: A report of the reading bands. Paper prepared for the annual meeting of the Australian Association for Research in Education, Armidale: NOvember, 1988.

Griffin, P. and Zbar V. (1988). *Assessment: Breaking New Ground. Working paper Number 1.* Monitoring Schools Project, Melbourne: State Board of Education.

Ingram, D. (1984). *Australian Second Language Proficiency Ratings.* Canberra : Commonwealth of Australia.

Liskin-Gasparro, J.G. (1984). The ACTFL proficiency guidelines: Gateways to testing and curriculum. *Foreign Language Annual.*

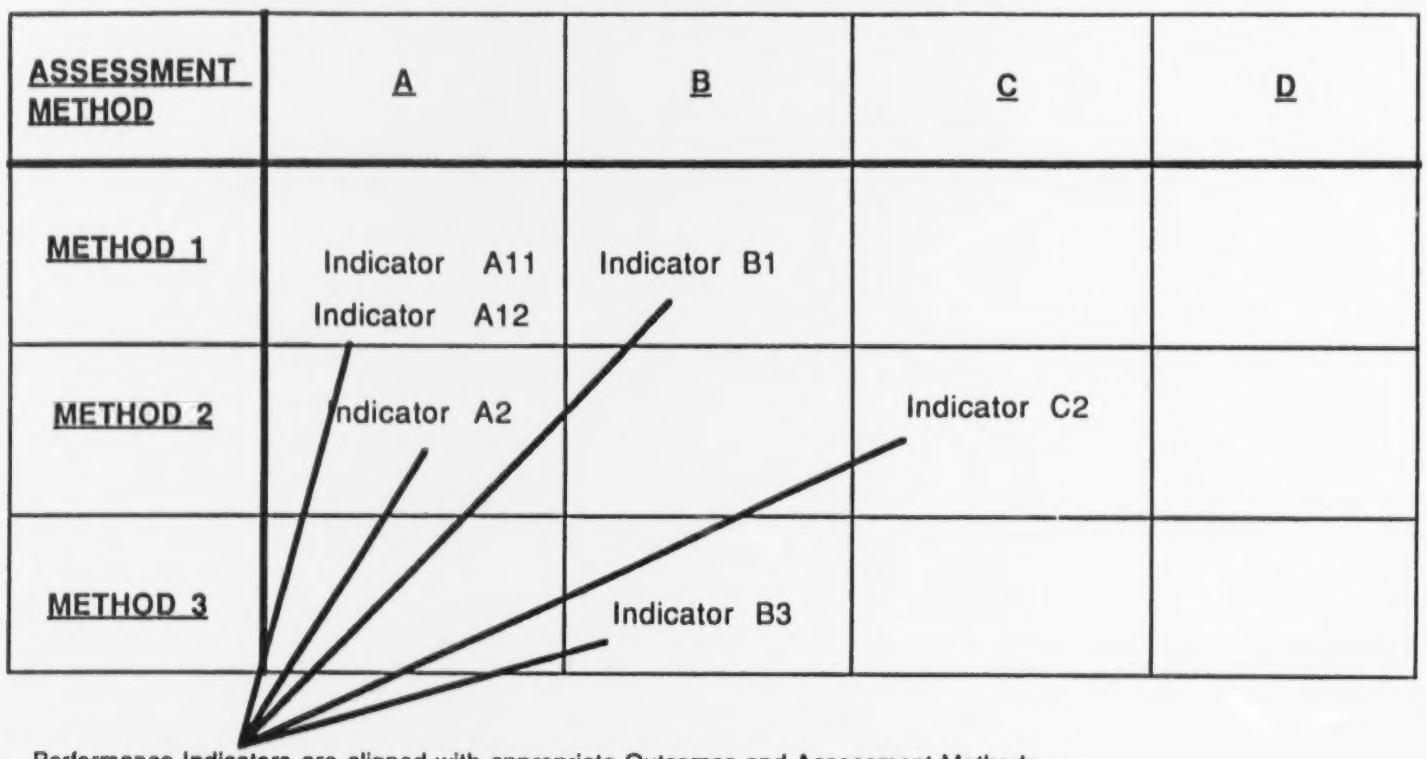
Mossensen, L., Hill, P. and Masters, G. (1987). *Test of Reading Comprehension.* Melbourne: ACER.

BEST COPY AVAILABLE

GOAL _____

OUTCOMES /ATTRIBUTES

ASSESSMENT METHOD	A	B	C	D
METHOD 1	Indicator A11 Indicator A12	Indicator B1		
METHOD 2	Indicator A2		Indicator C2	
METHOD 3		Indicator B3		



Performance Indicators are aligned with appropriate Outcomes and Assessment Methods

Figure 3

Matrix Worksheet for Matching Outcomesd, Assessment Methods and Indicators

MILESTONES

ASSESSMENT TECHNIQUES	ASKS OTHERS	USES VISUAL CUES	USES AUDITORY CUES	USES CONTEXT CUES
OBSERVATION & ANNECDOTAL RECORDS	asks adults/peers what a word is. asks adults/peers what a word means.	Eye moves between words and pictures. substitutes a similarly shaped word for the unknown word.	reuses words already heard in stories, wall stories or oral reading activities. uses first sound of a word when attempting a new word. Attempts to sound parts of words.	rereads sentence whenunable to read a word. expresses that it was the context that gave clues for the word. queries meaning of sentence whenunable to read word.
RUNNING RECORDS	asks adults/peers for meaning and pronunciationof a word.	uses appropriate substitutions e.g. house/ home.		
PARENT/ TEACHER CONFER-ENCE	asks parent for meaning and pronunciation of words.	states the picture helped to read the text.	states a word is known because it sounds right.	rereads sentence when unable to read a word. explains that it was the context that gave clues for the word.

Figure 4

Matrix Worksheet : Approach to Unknown Words